Statistical process control for serially correlated data

Wieringa, Jakob Edo

Chapter 6

Two worked out examples

In this chapter, we illustrate the use of control charts that were discussed in the previous chapters by two examples. The first is a real life example, based on a data set that appeared in Shewhart (1931). In his treatment of the data set, Shewhart did not take the presence of serial correlation into account. By using control charts of the previous chapters, we arrive at other conclusions than Shewhart did. The second example is a based on a simulated AR(1) series with a persistent change in the mean of the observations.

6.1 A real life example

The first book on quality control stems from the year 1931. It was written by the developer of the control chart: Dr. Walter A. Shewhart. In this very well written work, the newly developed concepts of quality control are illustrated with real-life examples. The second data set that appears in this book consists of 204 observations of the electrical resistance of a certain insulation material. In Table 6.1, these observations are reprinted in chronological order.

Shewhart decides to take subgroups of size four, and presents a control chart for the mean. The 51 subgroup averages are compared to control limits 'within which experience has shown that these observations should fall'. Since this data set is presented in one of the first chapters, Shewhart does not explain precisely how the control limits are computed. However, on page 296, Shewhart suggests computing the control limits as

$$\bar{\bar{X}} \pm 3\frac{\hat{\sigma}}{\sqrt{n}},$$

which is a formula that is very familiar to most SPC practitioners. The central line is the overall mean, which can be computed as 4,498MΩ. Estimating σ as the mean of the sample standard deviations of the subgroups, corrected by the well-known constant $c_4(4)$ to remove the bias (see for example Montgomery (1996), Table VI), results in a lower control limit of 4.006MΩ, and an upper control limit of 4,991MΩ. These values agree closely with the control limits that Shewhart depicted in the corresponding control chart. Eight of the subgroup averages fall outside the control limits, see Figure 6.2(a). Shewhart interprets these out-of-control signals as 'an indication of the existence of causes of variability which could be found and eliminated'. He reports that further research was instituted to find these causes of variability. The search was successful and a second control chart is presented, in which data points are depicted that were taken after elimination of these causes. All values remain within much tighter limits, and Shewhart concludes that 'this variation should be left to chance'.

Table 6.1: Electrical resistance of insulation in megohms

5,045	4,635	4,700	4,650	4,640	3,940	4,570	4,560	4,450	4,500	5,075	4,500
4,350	5,100	4,600	4,170	4,335	3,700	4,570	3,075	4,450	4,770	4,925	4,850
4,350	5,450	4,110	4,255	5,000	3,650	4,855	2,965	4,850	5,150	5,075	4,930
3,975	4,635	4,410	4,170	4,615	4,445	4,160	4,080	4,450	4,850	4,925	4,700
4,290	4,720	4,180	4,375	4,215	4,000	4,325	4,080	3,635	4,700	5,250	4,890
4,430	4,810	4,790	4,175	4,275	4,845	4,125	4,425	3,635	5,000	4,915	4,625
4,485	4,565	4,790	4,550	4,275	5,000	4,100	4,300	3,635	5,000	5,600	4,425
4,285	4,410	4,340	4,450	5,000	4,560	4,340	4,430	3,900	5,000	5,075	4,135
3,980	4,065	4,895	2,855	4,615	4,700	4,575	4,840	4,340	4,700	4,450	4,190
3,925	4,565	5,750	2,920	4,735	4,310	3,875	4,840	4,340	4,500	4,215	4,080
3,645	5,190	4,740	4,375	4,215	4,310	4,050	4,310	3,665	4,840	4,325	3,690
3,760	4,725	5,000	4,375	4,700	5,000	4,050	4,185	3,775	5,075	4,665	5,050
3,300	4,640	4,895	4,355	4,700	4,575	4,685	4,570	5,000	5,000	4,615	4,625
3,685	4,640	4,255	4,090	4,700	4,700	4,685	4,700	4,850	4,770	4,615	5,150
3,463	4,895	4,170	5,000	4,700	4,430	4,430	4,440	4,775	4,570	4,500	5,250
5,200	4,790	3,850	4,335	4,095	4,850	4,300	4,850	4,500	4,925	4,765	5,000
5,100	4,845	4,445	5,000	4,095	4,850	4,690	4,125	4,770	4,775	4,500	5,000

Source: Shewhart (1931), page 20, Table 2.
Reprinted with permission of the American Society for Quality.

However, if we take a closer look at the data set in Table 6.1, it appears that the successive values exhibit serial correlation. In fact, the data set appears to be a typical example of observations that can be successfully modelled using an AR(1) model. The sample autocorrelation function is exponentially declining, and the sample partial autocorrelation function shows a single spike at lag 1, see Figure 6.1.

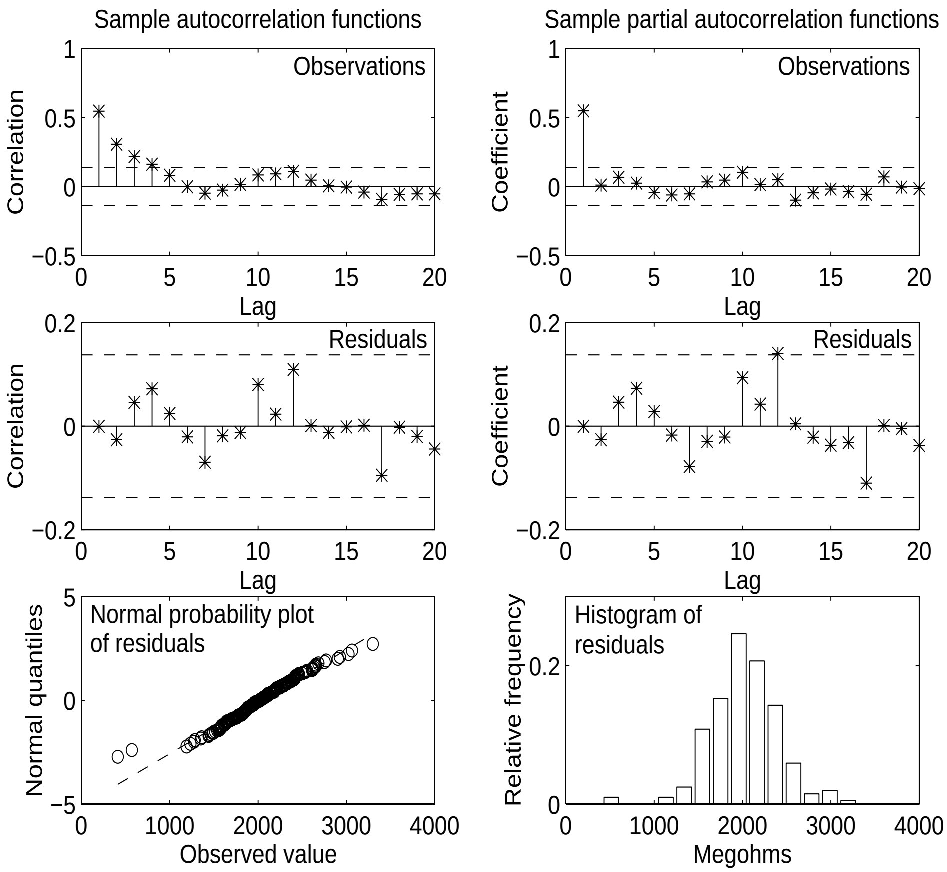

Figure 6.1: Analysis of the data in Table 6.1.

The AR parameter ϕ can be estimated as $\hat{\phi} = 0.549$. In Figure 6.1 it is shown that the residuals of this model show no significant serial correlation. A normal probability plot of the residuals indicates the presence of two or three outliers. All other observations lie more or less on a straight line, even after removal of the suspected data points. The histogram of the residuals is a little skewed to the right. Nevertheless, in our opinion, it is reasonable to assume that the residuals of this model are uncorrelated and normally distributed.

The autocorrelation function of the 51 subgroup averages shows less convincingly that serial correlation is present. Here we observe a well-known phenomenon (see also Section 7.3): by taking subgroup averages, the serial correlation is reduced. Ignoring the serial correlation in the data seems therefore justifiable. However, Wardell, Moskowitz, and Plante (1992) warn against this kind of taking of subgroups:

> "*However, if the data are truly autocorrelated, the points on the Shewhart chart will still show runs which are essentially due to correlation resulting from common causes rather than any special cause*".

This statement is in agreement with Figure 6.2, where five control charts corresponding to the data in Table 6.1 are depicted. Figure 6.2(a) shows a control chart of subgroup averages, as proposed by Shewhart (1931). In Figures 6.2(b)–(d), a modified Shewhart control chart, a Shewhart chart of residuals, and a Shewhart chart of modified residuals are depicted, respectively. In Figure 6.2(e), an EWMA chart of modified residuals is presented. We have omitted other EWMA charts and CUSUM-type control charts, since their results turned out to be very similar to the results of the EWMA chart of modified residuals.

The control limits in Figures 6.2(b)–(e) are adjusted so as to have an approximate in-control ARL of 370.4. This was presumably Shewhart's intention.

When comparing the width of the control limits of Figures 6.2(a)–(d) (the four Shewhart-type control charts), we observe that the control limits of Figures 6.2(b)–(d) are much wider than the control limits of Figure 6.2(a). This is to be expected, since the control chart of Figure 6.2(a) is constructed for monitoring subgroup averages, whereas the control charts in Figures 6.2(b)–(d) monitor statistics of individual observations.

However, the estimator of σ_Y that is used for Figure 6.2(a) is biased in case of small subsamples of serially correlated data, so that the width of the control limits is not correctly determined. The problem encountered here is similar to the problems of Sections 3.3, 4.2 and 5.2, concerning the bias in $\overline{MR}/d_2(2)$ as an estimator of σ_Y.

The control limits in Figure 6.2(a) are computed using the mean of the subgroup standard deviations divided by $c_4(4)$ as an estimator for σ_Y. In Section 7.5 of the next chapter, it is established that in case of serially correlated data, S^2 is a biased estimator for σ_Y^2. Using these results, it can be shown that the mean of the subgroup standard deviations underestimates σ_Y in case of positive autocorrelation. This bias disappears for large subgroup sizes. In the present case, where we have $n = 4$ and $\hat{\phi} = 0.549$, σ_Y is

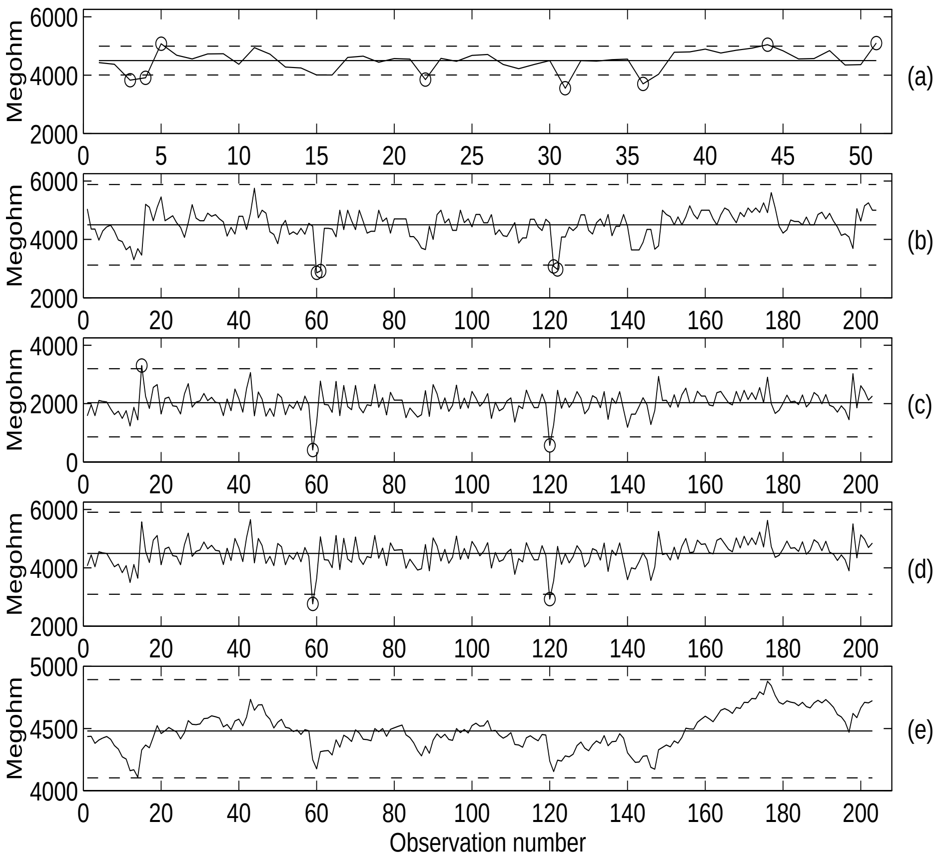

Figure 6.2: Five control charts corresponding to Table 6.1.
(a)=Chart for subgroup averages,
(b)=modified Shewhart,
(c)=Shewhart chart of residuals,
(d)=Shewhart chart of modified residuals,
(e)=EWMA chart of modified residuals.

underestimated, which results in control limits that are too tight. Consequently, too many false out-of-control signals are generated in the control chart for subgroup averages of Figure 6.2(a).

The Shewhart-type control charts of Figures 6.2(b)–(d) all show fewer out-of-control signals. Around observations 60 and 121, the presence of a special cause of variation may be suspected. Note that the modified Shewhart chart generates two signals around both of these observation numbers, whereas the Shewhart chart of residuals and the Shewhart chart of modified residuals produce only one out-of-control signal at both of these points. This illustrates the fact that the probability of an out-of-control signal on residuals-based control charts is high at the first observation following a shift in the mean, whereas this probability is much smaller at later observations, see also Section 3.5. Therefore, it may be wise to add observations 61 and 122 to the list of possible out-of-control situations. Based on an out-of-control signal on the Shewhart chart of residuals (Figure 6.2(c)), we have reason to suspect observation 16, too.

The EWMA control chart of modified residuals of Figure 6.2(e) does not produce any out-of-control signals, whereas the Shewhart-type control charts in Figures 6.2(b)–6.2(d) all signal shifts in the mean at the same observations. However, these out-of-control signals seem to relate to 'spike'-shifts in the mean. Shewhart-type control charts are to be preferred for this kind of shifts, since on these charts only the current observation is monitored. On EWMA charts and CUSUM charts, the effect of a 'spike'-shift is smoothed out, so that these charts are insensitive to changes in the mean of short duration, see also Lin and Adams (1996). However, as we will see in the following example, the EWMA of modified residuals is much more efficient in detecting persistent step changes in the mean than Shewhart-type control charts.

6.2 A simulated example

In this section, we will illustrate the use of the control charts that were considered in previous chapters in the case of a persistent step change in the mean of AR(1) observations. To this end, we simulated 150 observations of an AR(1) process with $\phi = 0.6$. The expectation of the first 79 observations is 2; at observation 80, a shift to $\mu = 3.25$ is introduced, which corresponds to a shift in the mean of the process of $1\sigma_Y$. In Figure 6.3, a modified Shewhart chart, a Shewhart chart of residuals, a Shewhart chart of modified residuals and an EWMA chart of modified residuals are used to monitor these observations.

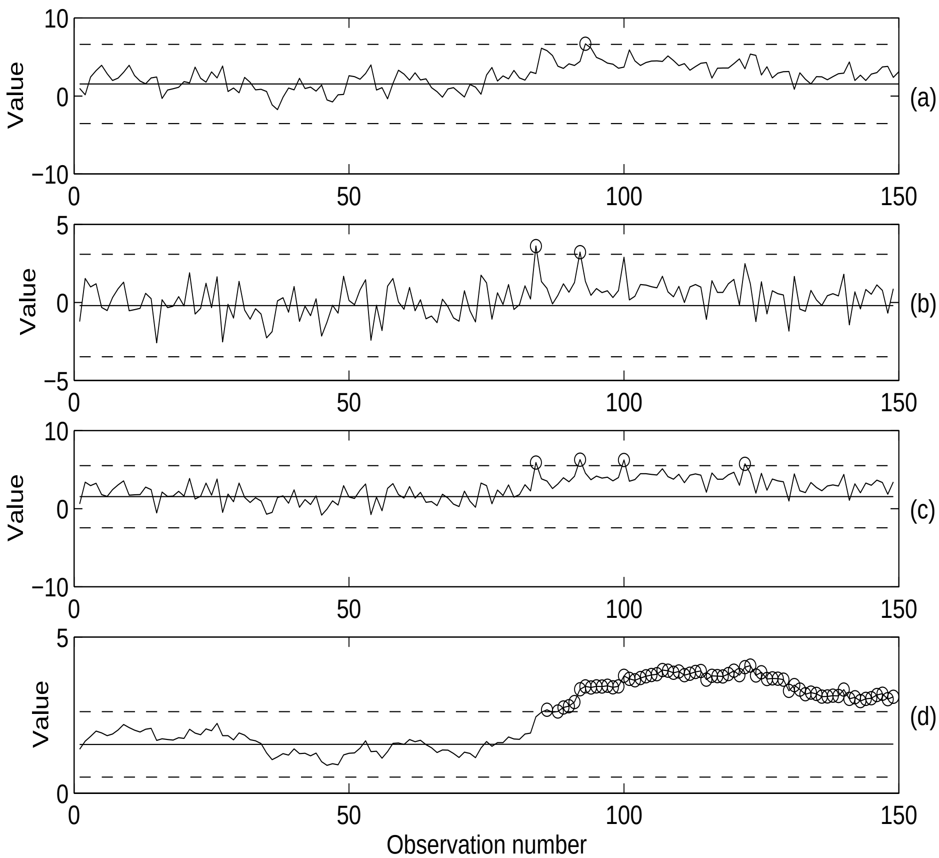

Figure 6.3: Control charts for a simulated AR(1) series.
(a)=modified Shewhart,
(b)=Shewhart chart of residuals,
(c)=Shewhart chart of modified residuals,
(d)=EWMA chart of modified residuals.

Figure 6.3(a) shows that the modified Shewhart chart signals at observation 93. The Shewhart chart of residuals in Figure 6.3(b) signals at the 84th and the 92nd residual. For the Shewhart chart of modified residuals in Figure 6.3(c) a value of $\lambda = 0.1$ is used for the EWMA. Out-of-control signals are observed at 84th, the 92nd, 100th, and the 122nd residual. The EWMA chart of modified residuals signals at observation 87, and every observation thereafter. The control limits are in all cases chosen such that